

**IMPLEMENTASI *SIMILARITY CHECK* PADA ABSTRAK SKRIPSI
MENGUNAKAN ALGORITMA *RABIN-KARP*
(STUDI KASUS PRODI S1 SEJARAH
FAKULTAS ILMU BUDAYA UNDIP)**



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
Pada Departemen Ilmu Komputer/ Informatika**

**Disusun oleh:
NADIA PERMATASARI ARUBEN
2401031240086**

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2019

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Nadia Permatasari Aruben

NIM : 24010312140086

Judul : Implementasi *Similarity Check* pada Abstrak Skripsi Menggunakan Algoritma
Rabin-Karp (Studi Kasus Prodi S1 Sejarah Fakultas Ilmu Budaya UNDIP)

Dengan ini saya menyatakan bahwa dalam tugas akhir/ skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 1 Agustus 2019



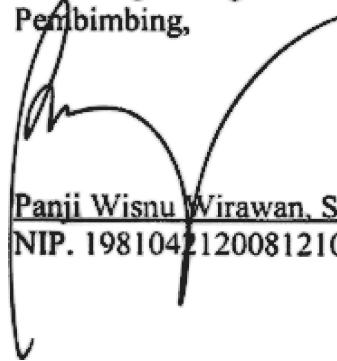
Nadia Permatasari Aruben
NIM. 24010312140086

HALAMAN PENGESAHAN

Judul : Implementasi *Similarity Check* pada Abstrak Skripsi Menggunakan Algoritma
Rabin-Karp (Studi Kasus Prodi S1 Sejarah Fakultas Ilmu Budaya Undip)
Nama : Nadia Permatasari Aruben
NIM : 24010312140086

Telah diujikan pada sidang tugas akhir tanggal 1 Agustus 2019

Semarang, 27 Agustus 2019
Pembimbing,



Panji Wisnu Wirawan, ST, MT
NIP. 198104212008121002

HALAMAN PENGESAHAN

Judul : Implementasi *Similarity Check* pada Abstrak Skripsi Menggunakan Algoritma *Rabin-Karp* (Studi Kasus Prodi S1 Sejarah Fakultas Ilmu Budaya Undip)

Nama : Nadia Permatasari Aruben

NIM : 24010312140086

Telah diujikan pada sidang tugas akhir tanggal 1 Agustus 2019

Semarang, 27 Agustus 2019

Mengetahui,

a.n. Ketua Departemen Ilmu Komputer/ Informatika FSM



Dr. Erna A. Wibowo, S.Si., M.Kom.

NIP. 198203092006041002

Panitia Penguji Skripsi,

Ketua,

Dr. Retno Kusumaningrum, S.Si., M.Kom.

NIP. 198104202005012001

ABSTRAK

Plagiarisme biasa terjadi di kalangan mahasiswa yang sedang mengerjakan tugas akhir, baik disengaja maupun tidak. Rendahnya etika akademik dan kurangnya pemahaman tentang plagiarisme menjadi alasan kuat hal tersebut masih sering terjadi. Oleh karena itu diperlukan proses untuk mendeteksi apakah skripsi mahasiswa mengandung plagiarisme. Salah satu bentuk plagiarisme adalah kemiripan kata di luar batas toleransi. Pendeteksian kemiripan kata pada skripsi mahasiswa tidak mungkin dilakukan secara manual mengingat banyaknya jumlah mahasiswa yang mengerjakan tugas akhir setiap semester. Untuk itu perlu dibangun sebuah sistem yang dapat mempermudah pendeteksian kemiripan kata, yaitu sistem *Similarity Check*. Implementasi *Similarity Check* pada penelitian ini menggunakan metode Pencocokan *String* dengan Algoritma *Rabin-Karp* untuk mendeteksi kemiripan kata pada sebuah dokumen abstrak skripsi terhadap korpus data abstrak skripsi. Algoritma ini bekerja dengan cara mencocokkan nilai *hash* dokumen uji yang telah dikalkulasi melalui *text preprocessing*, *parsing k-gram* dengan nilai *k* sebagai panjang *substring*nya, dan *rolling hash* untuk didapatkan nilai *hash*nya. Penelitian ini menggunakan 115 data abstrak skripsi yang dibagi menjadi 5 data sebagai data uji dan 110 data sebagai data pembandingan. Waktu pemrosesan (*runtime*) yang dihasilkan dari pengujian performa sistem *Similarity Check* sangat bergantung pada jumlah data pembandingan dan banyaknya kata penyusun keseluruhan data abstrak skripsi. Dalam pengujian performa dengan variasi jumlah data pembandingan dari 20 hingga 110 data dihasilkan rata-rata kenaikan *runtime* sebesar 1.52 detik. Pengujian akurasi sistem *Similarity Check* menghasilkan rata-rata nilai keakuratan sebesar 97,56%.

Kata kunci: plagiarisme, pencocokan *string*, algoritma *Rabin-Karp*, *similarity check* pada abstrak skripsi

ABSTRACT

Plagiarism is a common issue occurring among undergraduate students whom are undergoing final assignment to obtain their bachelor degree, regardless it is intentional or not. Many causes ranging from poor level of academic ethics to low level of awareness and knowledge became the reason for plagiarism. Therefore, a process of detecting content similarity is urgently required. Taking into account the amount of student conducting final assignment annually, detecting plagiarism manually is infeasible and need a system which able to detect words similarities. Implementation of Similarity Check System in this study used String Matching method with Rabin-Karp algorithm to detect the similarities between a document and a data corpus. This algorithm overall process consisted of matching the hash value of a document that had gone through (1) Text Preprocessing, (2) K-Gram Parsing with the value of k as its substring length, and (3) Rolling Hash to obtain its hash value. This study was conducted with 115 sample data of academic paper abstract of undergraduate programme in Departement of History, Faculty of Humanity, Diponegoro University which was separated into 5 testing samples and 110 comparison samples. Processing time of Similarity Check performance testing largely dependent upon the amount of comparison sample data used and the word count on all sample data, with the average increase of runtime as much as 1,52 seconds ranged from 20 to 110 comparison sample data. Meanwhile the accuracy achieved from accuracy testing is relatively high with the average accuracy rate of 97,56%.

keywords: plagiarism, string matching, Rabin-Karp algorithm, similarity check in academic paper abstract

KATA PENGANTAR

Segala puji syukur bagi Tuhan Yang Maha Agung atas kekuatan yang diberikan kepada penulis sehingga laporan tugas akhir yang berjudul “Implementasi Similarity Check pada Abstrak Skripsi Menggunakan Algoritma Rabin-Karp (Studi Kasus Prodi S1 Sejarah Fakultas Ilmu Budaya UNDIP)” dapat terselesaikan. Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan ini penulis banyak mendapat bimbingan dan bantuan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Dr. Retno Kusumaningrum, S.Si, M.Kom. selaku ketua Departemen Ilmu Komputer/ Informatika
2. Dr.Eng. Adi Wibowo, S.Si., M.T. selaku sekretaris Departemen Ilmu Komputer/ Informatika
3. Panji Wisnu Wirawan, ST, MT selaku Koordinator Tugas Akhir serta selaku dosen pembimbing yang telah memberikan waktu dan bimbingan yang tak ternilai bagi penulis
4. Emilia Widiyanti yang telah membantu dalam penyusunan logika sistem
5. Semua pihak yang telah membantu kelancaran dalam penyusunan tugas akhir, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan ini dapat bermanfaat bagi semuanya

Semarang, Agustus 2019

Penulis,

Nadia Permatasari Aruben

24010312140086

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Nadia Permatasari Aruben
NIM : 24010312140086
Program Studi : Informatika
Departemen : Ilmu Komputer/Informatika
Fakultas : Sains dan Matematika
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk **memberikan Hak Bebas Royalti Noneksklusif (Non-exclusive RoyaltyFree Right)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

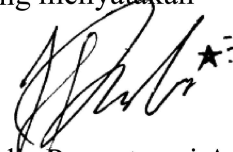
*Implementasi Similarity Check pada Abstrak Skripsi Menggunakan Algoritma Rabin-Karp
(Studi Kasus Prodi S1 Sejarah Fakultas Ilmu Budaya UNDIP)*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalihmedia/ formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 1 Agustus 2019

Yang menyatakan



Nadia Permatasari Aruben

24010312140086

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Tujuan dan Manfaat.....	3
1.4. Ruang Lingkup	3
1.5. Sistematika Penulisan.....	3
BAB II LANDASAN TEORI.....	5
2.1. Plagiarisme	5
2.2. Pencocokan <i>String</i> pada Data Teks	6
2.3. <i>Text Preprocessing</i>	7
2.4. <i>Parsing K-Gram</i>	9
2.5. Pencocokan <i>String</i> dengan Algoritma <i>Rabin-Karp</i>	10
2.5.1. <i>Rolling Hash</i>	11
2.5.2. Algoritma <i>Rabin-Karp</i>	12
2.6. Perhitungan Kemiripan.....	14
2.7. Model <i>Waterfall</i>	15
2.8. <i>Entity Relationship Diagram</i>	16
2.9. <i>Context Diagram</i>	17
2.10. <i>Data Flow Diagram</i>	17

2.11. <i>Flowchart</i>	18
2.12. PHP	19
2.13. MySQL	19
2.14. Pengukuran Kesalahan	19
BAB III SPESIFIKASI, ANALISIS, DAN PERANCANGAN	20
3.1. Proses <i>Similarity Check</i>	20
3.2. Spesifikasi dan Analisis Kebutuhan Sistem	25
3.2.1. Spesifikasi Sistem	25
3.2.2. Pemodelan Data	25
3.2.3. Pemodelan Fungsional	26
3.2.3.1. <i>Context Diagram</i>	26
3.2.3.2. <i>Decomposition Diagram</i>	27
3.2.3.3. DFD Level 1 Sistem <i>Similarity Check</i>	27
3.2.3.4. DFD Level 2 Sub Proses Mengelola Data Abstrak	29
3.2.3.5. DFD Level 2 Sub Proses <i>Preprocessing</i>	30
3.2.3.6. DFD Level 2 Sub Proses Melakukan Analisa Kemiripan	31
3.3. Perancangan Sistem	32
3.3.1. Perancangan Proses Sistem <i>Similarity Check</i>	32
3.3.1.1. <i>Text Preprocessing</i>	33
3.3.1.2. Proses <i>Parsing K-Gram</i>	36
3.3.1.3. Proses <i>Rolling Hash</i>	36
3.3.1.4. Proses Pencocokan <i>String</i> menggunakan Algoritma <i>Rabin-Karp</i> serta Melakukan Perhitungan Persentase Kemiripan	37
3.3.2. Perancangan Antarmuka Sistem <i>Similarity Check</i>	38
3.3.2.1. Perancangan Antarmuka Beranda	38
3.3.2.2. Perancangan Antarmuka Pilih Dokumen	39
3.3.2.3. Perancangan Antarmuka Tambah Data Abstrak	39
3.3.2.4. Perancangan Antarmuka Data Abstrak	40
3.3.2.5. Perancangan Antarmuka <i>Stopword</i>	40
3.3.2.6. Perancangan Antarmuka Hasil Analisa Kemiripan	41
3.3.2.7. Perancangan Antarmuka Riwayat Pengujian	41
BAB IV IMPLEMENTASI DAN PENGUJIAN	43

4.1. Implementasi	43
4.1.1. Spesifikasi Perangkat	43
4.1.2. Implementasi Basis Data	43
4.1.3. Implementasi Fungsi.....	44
4.1.3.1. Fungsi <i>Preprocessing</i>	44
4.1.3.2. Fungsi <i>Parsing K-Gram</i>	46
4.1.3.3. Fungsi <i>Rolling Hash</i>	46
4.1.3.4. Fungsi Pencocokan <i>String</i>	47
4.1.4. Implementasi Antarmuka Sistem <i>Similarity Check</i>	47
4.1.4.1. Implementasi Antarmuka Beranda	48
4.1.4.2. Implementasi Antarmuka Pilih Dokumen dan Hasil Analisa Kemiripan	48
4.1.4.3. Implementasi Antarmuka Tambah Data Abstrak	49
4.1.4.4. Implementasi Antarmuka Data Abstrak	50
4.1.4.5. Implementasi Antarmuka <i>Stopword</i>	51
4.1.4.6. Implementasi Antarmuka Riwayat Pengujian	51
4.2. Pengujian	52
4.2.1. Rencana Pengujian.....	52
4.2.2. Pelaksanaan Pengujian.....	52
4.2.3. Analisis Hasil Pengujian.....	53
BAB V PENUTUP.....	54
5.1. Kesimpulan.....	54
5.2. Saran	54
DAFTAR PUSTAKA.....	55
LAMPIRAN-LAMPIRAN	57

DAFTAR GAMBAR

Gambar 2. 1.	Proses <i>Case Folding</i>	7
Gambar 2. 2.	Proses <i>Filtering</i>	8
Gambar 2. 3.	Proses <i>Stemming</i>	9
Gambar 2. 4.	Proses <i>Parsing</i> Menggunakan 5-gram Menghasilkan 10 <i>Substring</i>	10
Gambar 2. 5.	<i>Fingerprint</i> Awal	12
Gambar 2. 6.	Pergeseran <i>Substring</i>	13
Gambar 2. 7.	Komparasi Kedua	13
Gambar 2. 8.	Komparasi Ketiga	13
Gambar 2. 9.	Komparasi Keempat dengan Nilai <i>Hash</i> Sama	13
Gambar 2. 10.	Komparasi Kelima dengan Nilai <i>String</i> Sama	14
Gambar 2. 11.	Model <i>Waterfall</i>	15
Gambar 3. 1.	Diagram Proses Sistem <i>Similarity Check</i>	20
Gambar 3. 2.	ERD Sistem <i>Similarity Check</i>	26
Gambar 3. 3.	<i>Context Diagram</i> Sistem <i>Similarity Check</i>	26
Gambar 3. 4.	<i>Decomposition Diagram</i> Sistem <i>Similarity Check</i>	27
Gambar 3. 5.	DFD Level 1 Sistem <i>Similarity Check</i>	28
Gambar 3. 6.	DFD Level 2 Sub Proses Mengelola Data Abstrak	29
Gambar 3. 7.	DFD Level 2 Sub Proses <i>Preprocessing</i>	30
Gambar 3. 8.	DFD Level 2 Sub Proses Melakukan Analisis Kemiripan	31
Gambar 3. 9.	<i>Flowchart</i> Sistem <i>Similarity Check</i>	33
Gambar 3. 10.	<i>Flowchart Preprocessing</i>	34
Gambar 3. 11.	<i>Flowchart</i> Proses <i>Case Folding</i>	34
Gambar 3. 12.	<i>Flowchart</i> Proses <i>Filtering</i>	35
Gambar 3. 13.	<i>Flowchart</i> Proses <i>Stemming</i>	35
Gambar 3. 14.	<i>Flowchart</i> Proses <i>Parsing K-Gram</i>	36
Gambar 3. 15.	<i>Flowchart</i> Proses <i>Rolling Hash</i>	37
Gambar 3. 16.	<i>Flowchart</i> Proses Pencocokan <i>String</i> Menggunakan Algoritma Rabin-Karp dan Perhitungan Presentase Kemiripan	38
Gambar 3. 17.	Perancangan Antarmuka Beranda	39
Gambar 3. 18.	Perancangan Antarmuka Pilih Dokumen	39

Gambar 3. 19.	Perancangan Antarmuka Tambah Data Abstrak	40
Gambar 3. 20.	Perancangan Antarmuka Data Abstrak.....	40
Gambar 3. 21.	Perancangan Antarmuka <i>Stopword</i>	41
Gambar 3. 22.	Perancangan Antarmuka Hasil Analisa Kemiripan	41
Gambar 3. 23.	Perancangan Antarmuka Riwayat Pengujian	42
Gambar 4. 1.	Struktur <i>Database</i> Sistem <i>Similarity Check</i>	43
Gambar 4. 2.	Implementasi Antarmuka Beranda	48
Gambar 4. 3.	Implementasi Antarmuka Pilih Dokumen	49
Gambar 4. 4.	Implementasi Antarmuka Hasil Analisa Kemiripan.....	49
Gambar 4. 5.	Implementasi Antarmuka Tambah Data Abstrak	50
Gambar 4. 6.	Implementasi Antarmuka Data Abstrak	50
Gambar 4. 7.	Implementasi Antarmuka <i>Stopword</i>	51
Gambar 4. 8.	Implementasi Antarmuka Riwayat Pengujian	51

DAFTAR TABEL

Tabel 2. 1.	Notasi ERD	17
Tabel 2. 2.	Notasi DFD	18
Tabel 2. 3.	Simbol-simbol <i>Flowchart</i>	18
Tabel 3. 1.	Contoh Masukan Dokumen Abstrak	21
Tabel 3. 2.	Hasil <i>Case Folding</i>	21
Tabel 3. 3.	Hasil <i>Filtering</i> Abstrak Korpus dan Abstrak Uji	22
Tabel 3. 4.	Hasil <i>Stemming</i> Abstrak Korpus dan Abstrak Uji	22
Tabel 3. 5.	Hasil <i>Parsing 5-gram</i> Dokumen Abstrak Korpus	22
Tabel 3. 6.	Hasil <i>Parsing 5-gram</i> Dokumen Abstrak Uji	22
Tabel 3. 7.	Nilai <i>ASCII</i> karakter dalam <i>substring</i> “lgori” dan “gorit”	23
Tabel 3. 8.	Hasil <i>Rolling Hash</i> Dokumen Abstrak Korpus	24
Tabel 3. 9.	Hasil <i>Rolling Hash</i> Dokumen Abstrak Uji	24
Tabel 3. 10.	Tabel Spesifikasi Kebutuhan Fungsional	25
Tabel 4. 1.	Struktur Tabel Abstrak	44
Tabel 4. 2.	Struktur Tabel Uji	44
Tabel 4. 3.	Struktur Tabel <i>Stopword</i>	44

BAB I

PENDAHULUAN

Bab pendahuluan menjelaskan mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, dan ruang lingkup.

1.1. Latar Belakang

Skripsi merupakan karya ilmiah yang harus dibuat oleh setiap mahasiswa yang ingin mendapatkan gelar sarjana pada perguruan tinggi. Banyaknya mahasiswa yang menyusun skripsi tiap semester memungkinkan terjadinya kasus plagiarisme baik yang disengaja maupun tidak disengaja dengan alasan yang bermacam-macam. Batasan antara plagiarisme dengan riset sendiri masih diragukan karena dalam sebuah penelitian harus ada rujukan sumber yang dapat dibuktikan kebenarannya agar dapat menyokong poin yang diajukan pada riset tersebut. Kasus ini menunjukkan bahwa plagiarisme yang dikelompokkan hanya berdasarkan perhitungan banyaknya kata yang sama dengan dokumen sumber adalah tidak valid nilainya (Maurer, Kappe, & Zaka, 2006).

Kasus plagiarisme dapat disebabkan banyak faktor, secara umum ada dua hal yang menjadi alasan seseorang melakukan plagiarisme, yang pertama karena tidak mengetahui bahwa tindakan tersebut merupakan plagiat. Kurangnya pemahaman tentang aturan penulisan karya ilmiah, kemampuan menulis yang rendah, dan kurang memahami materi dari topik skripsi menjadi alasan kuat mengapa plagiarisme dilakukan. Alasan lainnya adalah dengan tidak sengaja melakukan plagiarisme karena tanpa sadar menyalin sumber-sumber yang telah dibaca sebelumnya (Silvana, Rullyana, & Hadiapurwa, 2017). Kasus plagiarisme juga terjadi pada kampus FIB, mengingat jumlah mahasiswa yang cukup banyak per semesternya. Melakukan pengecekan secara manual satu per satu setiap skripsi yang dibuat oleh mahasiswa tidak memungkinkan karena jumlah skripsi yang harus diperiksa terus bertambah tiap semester dan membutuhkan waktu serta tenaga yang cukup besar sehingga diperlukan adanya perangkat lunak yang dapat membantu mempermudah pengecekan kemiripan konten pada skripsi. Tugas akhir ini menggunakan dokumen abstrak skripsi Prodi Sejarah untuk dibandingkan kemiripannya dengan pertimbangan skripsi Prodi Sejarah

tidak ada yang sama karena tiap skripsi harus mengusulkan permasalahan yang berbeda, jadi dapat dipastikan secara konten tidak mungkin dilakukan plagiarisme *word by word* atau *copy-paste* kalimat melainkan lebih condong ke plagiarisme ide dan *style* penulisan pada skripsi yang membahas topik yang serupa. Peran perangkat lunak di sini adalah membantu mendeteksi dugaan adanya plagiarisme dari tingkat kemiripan abstraknya, suatu skripsi yang membahas topik mengenai kehidupan sosial ekonomi di desa A dan desa B tidak mungkin memiliki konten yang sama, namun bisa saja *style* penulisan banyak yang mirip, kemungkinan kemiripan konten dari skripsi tersebut dapat dideteksi hanya dari membandingkan isi abstraknya.

Upaya yang dapat dilakukan untuk mendeteksi plagiarisme salah satunya dengan dilakukan pendeteksian kemiripan dokumen dengan bantuan perangkat lunak (Wibowo, 2012). Pendeteksian kemiripan kata menggunakan perangkat lunak masih harus disertai dengan pertimbangan manual apakah kata-kata yang terdeteksi sama benar merupakan suatu plagiarisme. Hal ini dikarenakan perangkat lunak hanya memberikan keluaran berupa informasi persentase kemiripan dengan mendeteksi kemunculan kata yang sama pada kedua dokumen yang diujikan tanpa mempertimbangkan apakah keseluruhan kalimat bermakna sama, maka dari itu dibutuhkan pengecekan secara manual pada dokumen yang terindikasi mengandung plagiarisme.

Dalam membangun perangkat lunak pendeteksi kemiripan kata dapat digunakan metode pencocokan *string* dengan beberapa macam algoritma seperti algoritma *Rabin-Karp*, algoritma *Jaro-Winkler Distance*, dan algoritma *Knut-Morris-Pratt*. Masing-masing algoritma memiliki kelebihan dan kekurangan tersendiri. Penelitian Leonardo & Hansun (2017) melakukan perbandingan hasil perhitungan menggunakan algoritma *Rabin-Karp* dengan algoritma *Jaro-Winkler Distance*, didapatkan kesimpulan algoritma *Rabin-Karp* memiliki tingkat akurasi dan kecepatan yang lebih tinggi pada pencarian *string* dalam dokumen besar sedangkan algoritma *Jaro-Winkler Distance* memberikan hasil yang cukup baik hanya pada *string* pendek. Penelitian lain oleh Andres, Christopher, & Saloko (2006) membandingkan algoritma *Rabin-Karp* dengan algoritma *Knut-Morris-Pratt* menghasilkan kesimpulan algoritma *Rabin-Karp* memiliki keunggulan dalam memproses *string* panjang dan penggunaan memori dalam algoritma ini lebih efektif dibanding algoritma *Knut-Morris-Pratt*.

Berdasarkan uraian di atas, pada penelitian tugas akhir ini akan dilakukan penerapan *similarity check* pada abstrak skripsi menggunakan algoritma *Rabin-Karp* dengan studi kasus pada Prodi S1 Sejarah FIB UNDIP.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas dapat dirumuskan permasalahan yang dapat diambil yaitu, bagaimana melakukan *similarity check* pada abstrak skripsi menggunakan algoritma *Rabin-Karp* dengan studi kasus pada Prodi S1 Sejarah FIB UNDIP.

1.3. Tujuan dan Manfaat

Tujuan dari penelitian tugas akhir ini adalah melakukan implementasi *similarity check* pada abstrak skripsi menggunakan algoritma *Rabin-Karp* dengan studi kasus pada Prodi S1 Sejarah FIB UNDIP.

Manfaat dari penelitian tugas akhir ini adalah mempermudah pengecekan kemiripan kata pada abstrak skripsi dengan hasil penelitian yang telah ada sebelumnya.

1.4. Ruang Lingkup

Ruang lingkup dari tugas akhir ini adalah sebagai berikut.

1. Dokumen *input* berupa data abstrak berbentuk *text file* yang sebelumnya telah dimasukkan ke dalam *database*.
2. Data *text file* bersifat *plain text*, yaitu *file* teks berupa huruf dan angka tanpa mencakup gambar, tabel, dan sejenisnya dengan bahasa penulisan yang digunakan adalah bahasa Indonesia.
3. *Similarity Check* dilakukan antara satu dokumen abstrak uji terhadap korpus data abstrak. Data abstrak terdiri dari skripsi yang telah tersimpan dalam Perpustakaan Departemen Sejarah.
4. Penelitian diimplementasikan berbasis *web*, menggunakan bahasa pemrograman PHP dan sistem basis data MySQL.
5. Sistem ini dirancang dengan menggunakan model *Waterfall*, yang akan dilakukan hanya sampai pada tahap implementasi dan pengujian unit.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan dalam pembuatan tugas akhir mengenai implementasi *similarity check* pada abstrak skripsi menggunakan algoritma *Rabin-Karp*.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan tinjauan pustaka yang berhubungan dengan topik tugas akhir berupa penelitian yang terkait dengan, plagiarisme, *text preprocessing*, *parsing k-gram*, algoritma *rabin-karp*, *waterfall*, PHP, dan MySQL.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan tahapan yang dilakukan pada penelitian tugas akhir. Pada bab ini akan dijelaskan mengenai gambaran umum proses *similarity check* beserta perhitungan manual proses pencocokan *string*, spesifikasi dan analisis kebutuhan sistem serta desain sistem *similarity check*.

BAB IV HASIL DAN ANALISIS

Bab ini menyajikan pembahasan mengenai implementasi basis data, fungsi, dan antarmuka sistem serta pengujian performa dan akurasi sistem *similarity check*.

BAB V PENUTUP

Bab ini merupakan kesimpulan dari bab-bab yang dibahas sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.